Stochastic Processes and their Applications 21 (1986) 291-304
North-Holland

OPTIMAL FOURIER-HERMITE EXPANSION FOR ESTIMATION

James Ting-Ho LO

Department of Mathematics, University of Maryland Baltimore County, Catonsville, MD 21228, USA

Sze-Kui NG

Department of Mathematics, Hong Kong Baptist College, 224 Waterloo Road, Kowloon, Hong Kong

Received 27 February 1984
Revised 9 September 1985

The purpose of the paper is to present a systematic method for developing an approximate *recursive* estimator which is optimal for the given structure and approaches the best estimate, when the order of approximation increases.

The minimal variance estimate is projected onto the Hilbert subspace of all Fourier-Hermite (FH) series, driven by the observations, with the same given index set. The projection results in a system of linear algebraic equations for the FH coefficients, the parameters of the desired approximate estimator.

The estimator consists of finitely many Wiener integrals of the observations and a memoryless nonlinear postprocessor. The postprocessor is an arithmetic combination of the Hermite polynomials evaluated at the Wiener integrals. A couple of recursive methods for calculating the Wiener integrals are included.

recursive estimation * minimum variance * Fourier-Hermite expansions * Wiener integrals

1. Introduction

The Volterra series [3] and the series of multiple Wiener integrals [1-2] were used to approximate the minimum variance (MV) filter by Katzenelson and Gould [5] and Mitter and Ocone [6-8] respectively. Their results suffer from the following two difficulties.

(1) The integrands of the multiple (or iterated) integrals in their series expansions are characterized by the nth order Wiener-Hopf equations, which are hard to solve.

(2) The multiple integrals in their series expansions are hard to evaluate. There is not a recursive algorithm for their evaluation.

In this paper, we shall show that these difficulties can be avoided by the use of the Fourier-Hermite (FH) functionals developed by Cameron and Martin [4].

This work was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-80-0241.

0304-4149/86/$3.50

Let $\{\phi_i\}$ be a complete orthonormal set (CONS) of functions on $[0, t]$. A finite FH series driven by the observation $y^t := \{y_\tau, 0 \leq \tau \leq t\}$ is

$$b(t,0) + \sum_{k=1}^{n} \sum_{\xi \subset \underset{\sim}{\xi}_k} b(t,k,\xi) G_k(t, \phi_{\xi_1} \cdots \phi_{\xi_k}, y)$$

where G_k denotes the multiple Wiener integral of $\phi_{\xi_1} \cdots \phi_{\xi_k}$ wrt y [1, 4, 9] and $\underset{\sim}{\xi}_k$ is a finite set of k-tuples $\xi = [\xi_1, \ldots, \xi_k]$ of nonnegative integers. The multiple Wiener integrals in the above series are called FH functionals.

The FH functionals are orthonormal with respect to the Wiener measure. This property enables us to project the minimum variance estimate $\hat{\Phi}(x)$ of a measurable functional Φ of the signal process x, such that $E\Phi^2(x) < \infty$, directly onto Hilbert subspaces $H_n(t, \underset{\sim}{\xi})$ of all finite FH series with the same index set $\underset{\sim}{\xi} = \underset{\sim}{\xi}_1 \cup \cdots \cup \underset{\sim}{\xi}_k$. A typical element in $H_n(t, \underset{\sim}{\xi})$ is the FH series displayed above. The direct projection for a preassigned index set does not need the nth order Wiener–Hopf equation. In fact, it produces a system of linear algebraic equations for the FH coefficients $b(t, k, \xi)$!

By expanding $b(t, k, \xi)$ into finite Fourier series on the time interval $[0, T]$, over which the estimator is applied, we obtain a further and simpler approximate estimator. The Fourier expansion is carried out by minimizing a weighted cumulative estimation error over $[0, T]$.

A most remarkable property of an FH functional is that it is the product of finitely many Hermite polynomials. The argument of each Hermite polynoimal is simply a Wiener integral $\int_0^t \phi_{\xi_i}(t, s_i)\, \mathrm{d}y(s_i)$! A couple of recursive methods to evaluate the Wiener integral are given at the end of this paper.

2. Problem statement and preliminaries

In this paper, we mainly consider the estimation problem for the scalar system that is described by the following signal and sensor equations in the sense of Ito:

$$\mathrm{d}x = f(x,t)\,\mathrm{d}t + g(x,t)\,\mathrm{d}v, \qquad \mathrm{d}y = h(x,t)\,\mathrm{d}t + r_t^{1/2}\,\mathrm{d}w \tag{1}$$

where w and v are independent standard Wiener processes and x_0 is an independent random variable with a given distribution [11–13]. It is assumed that the functions f, g, and h are such that the signal equation has a unique solution and the measures μ_y and μ_w induced by y and w respectively on the measurable space $(C[0, t], B)$ of the continuous functions and the associated Borel field are equivalent and, almost surely with respect to u_y,

$$\frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_w}(y) = E_{\mu_x}(\theta(x,y)) > 0,$$

$$\theta^t := \theta^t(x,y) := \exp\left[\int_0^t h_\tau r_\tau^{-1}\,\mathrm{d}y_\tau - \tfrac{1}{2}\int_0^t h_\tau^2 r_\tau^{-1}\,\mathrm{d}\tau\right]$$

$$h_t := h(x_t, t).$$

Furthermore the conditional expectation $\hat{\Phi}(x) := E(\Phi(x)|y^t)$ of a measurable functional $\Phi(x)$ of x, such that $E\Phi^2(x) < \infty$, given the measurement $y^t := \{y_\tau, 0 \le \tau \le t\}$ has the following representation [11-13]:

$$\hat{\Phi}(x) = \sigma_t(\Phi)/\sigma_t(1), \qquad \sigma_t(\Phi) := (\sigma_t(\Phi))(y) := E_{\mu_x}(\Phi(x)\theta^t(x, y))$$

where 1 is the constant function with value one everywhere. The purpose of this paper is to expand $\hat{\Phi}(x)$ into a finite Fourier-Hermite series.

Throughout the paper, we will adopt the following symbols:

$$\int^k (\ )\,\mathrm{d}^k w := \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} (\ )\,\mathrm{d}w_{s_k} \cdots \mathrm{d}w_{s_1},$$

$$\fint^k (\ )\,\mathrm{d}^k \sigma := \int_0^t \cdots \int_0^t (\ )\,\mathrm{d}\sigma_k \cdots \mathrm{d}\sigma_1.$$

We note that θ^t satisfies

$$\mathrm{d}\theta^t = h_t r_t^{-1} \theta^t\,\mathrm{d}y_t.$$

Hence by the Peano-Baker procedure, it has the following finite expansion in iterated integrals:

$$\begin{aligned}\theta^t &= \sum_{k=0}^{n} \int^k (h_{s_1} r_{s_1}^{-1}) \cdots (h_{s_k} r_{s_k}^{-1})\,\mathrm{d}^k y \\ &\quad + \int^{n+1} (h_{s_1} r_{s_1}^{-1}) \cdots (h_{s_{n+1}} r_{s_{n+1}}^{-1}) \theta^{s_{n+1}}\,\mathrm{d}^{n+1} y,\\ \sigma_t(\Phi) &= \sum_{k=0}^{n} \int^k E[(h_{s_1} r_{s_1}^{-1}) \cdots (h_{s_k} r_{s_k}^{-1})\Phi(x)]\,\mathrm{d}^k y \\ &\quad + \int^{n+1} E_{\mu_x}[(h_{s_1} r_{s_1}^{-1}) \cdots (h_{s_{n+1}} r_{s_{n+1}}^{-1})\Phi(x)\theta^{s_{n+1}}]\,\mathrm{d}^{n+1} y. \end{aligned} \tag{3}$$

A product-to-sum formula for iterated integrals with respect to the Wiener process w was found by Shigekawa [15, p. 276], which is necessary in our orthogonal expansion. We shall now restate it as a theorem.

Theorem 2.1. *Let a and b be functions from $L^2([0, t]^n)$ and $L^2([0, t]^m)$ respectively. Then the product of two iterated stochastic integrals has finite iterated integral expansion as follows:*

$$\left[\int^n a(s_1, \ldots, s_n)\,\mathrm{d}^n w(s)\right]\left[\int^m b(s_1, \ldots, s_m)\,\mathrm{d}^m w(s)\right]$$

$$= \sum_{k=0}^{\min(n,m)} \frac{1}{k!} \int^{n+m-2k} \left[\fint^k \sum_{\pi} \bar{a}(\sigma_1, \ldots, \sigma_k, s_{\pi(1)}, \ldots, s_{\pi(n-k)}) \right.$$

$$\left. \bar{b}(\sigma_1, \ldots, \sigma_k, s_{\pi(n-k+1)}, \ldots, s_{\pi(n+m-2k)})\,\mathrm{d}^k\sigma\right] \mathrm{d}^{n+m-2k} w(s), \tag{4}$$

$$\bar{a}(s_1, \ldots, s_n) := a(s_{\tau(1)}, \ldots, s_{\tau(n)}), \tag{5}$$

$\tau :=$ *a permutation of* $1, \ldots, n$ *such that* $s_{\tau(1)} \geq s_{\tau(2)} \geq \cdots \geq s_{\tau(n)}$,

$$\mathrm{d}^k\sigma := \mathrm{d}\sigma_k \cdots \mathrm{d}\sigma_1 \tag{6}$$

where the summation $\sum_\pi$ is taken over all combinations π of $n+m-2k$ elements $\{1, \ldots, n+m-2k\}$ taken $n-k$ at a time.

The multiple Wiener integral $G_k(t, a, w)$ of $a(s_1, \ldots, s_k)$ [1-2] and the iterated integral $\int^k a \, \mathrm{d}^k w$ are related as follows:

$$G_k(t, a, w) = k! \int^k \tilde{a}(s_1, \ldots, s_k) \, \mathrm{d}^k w, \tag{7}$$

$$\int^k a(s_1, \ldots, s_k) \, \mathrm{d}^k w = \frac{1}{k!} G_k(t, \bar{a}, w) \tag{8}$$

where $\bar{a}$ is defined in (5) and $\tilde{a}$ denotes the symmetrization of a.

Let $\{\phi_i, i = 1, 2, \ldots\}$ be a CONS on $[0, t]$ and consider the multiple Wiener integral

$$F(t, k, \xi, w) := G_k(t, \phi_{\xi_1} \cdots \phi_{\xi_k}, w) \tag{9}$$

where ξ_i is a nonnegative integer for $i = 1, \ldots k$ and $\xi_1 \leqslant \cdots \leqslant \xi_k$. The set of all such integrals, which are orthonormal, was called the Fourier-Hermite (FH) set by Cameron and Martin [4]. It was shown in [4] that if $A(w')$ is any functional for which $E[A^2(w')] < \infty$, then it can be approximated arbitrarily close by a FH series in the form:

$$A(s') \simeq \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k}^{n} a(k; \xi) F(t, k, \xi, w), \qquad a(k; \xi) := E[A(w') F(t, k, \xi, w)] \tag{10}$$

where $\underset{\sim}{\xi}_k$ is a finite subset of $\{\xi = [\xi_1, \ldots, \xi_k] | \xi_1 \leqslant \cdots \leqslant \xi_k, \ \xi_i$ is a nonnegative integer for $i = 1, \ldots, k\}$.

It is necessary to impose the restriction $\xi_1 \leqslant \cdots \leqslant \xi_k$ for each index vector, ξ, in order to have orthonormal FH functionals associated with different vectors. For instance, $G(t, \phi_1\phi_2, w)$ and $G(t, \phi_2\phi_1, w)$ are identical

We shall need the following terminology:

(i) The integral (9) is called the ξth FH functional;

(ii) The term $a(k; \xi) F(t, k, \xi, w)$ is called the ξth FH component;

(iii) The coefficient $a(k; \xi)$ is called the ξth FH coefficient of $A(w')$;

(iv) The set $\underset{\sim}{\xi} := \underset{\sim}{\xi}_1 \cup \cdots \cup \underset{\sim}{\xi}_n$ is called the index set of the finite FH series on the right side of (10).

From (3), we have

$$\begin{aligned}
\sigma_t(\Phi) &= \alpha(t, 0) + \sum_{k=1}^{n} \int^k \alpha_k(t, s_1, \ldots, s_k) \, \mathrm{d}^k y \\
&\quad + \int^{n+1} r_{n+1}(t, s_1, \ldots, s_{n+1}, y^{s_{n+1}}) \, \mathrm{d}^{n+1} y, \\
\sigma_t(1) &= \beta(t, 0) + \sum_{k=1}^{2n} \int^k \beta_k(t, s_1, \ldots, s_k) \, \mathrm{d}^k y \\
&\quad + \int^{2n+1} r_{2n+1}(t, s_1, \ldots, s_{2n+1}, y^{s_{2n+1}}) \, \mathrm{d}^{2n+1} y.
\end{aligned} \tag{11}$$

Given index sets $\underset{\sim}{\xi} = \xi_1 \cup \cdots \cup \xi_n$ and $\underset{\sim}{\eta} = \eta_1 \cup \cdots \cup \eta_{2n}$, the finite FH series representations of $\sigma_t(\Phi)$ and $\sigma_t(1)$ can now be written as

$$\sigma_t(\Phi) = \alpha(t, 0) + \sum_{k=1}^{n} \sum_{\xi \in \xi_k} \alpha(t, k, \xi) F(t, k, \xi, y) + R_1(y', \underset{\sim}{\xi}),$$

$$\sigma_t(1) = \beta(t, 0) + \sum_{k=1}^{2n} \sum_{\eta \in \eta_k} \beta(t, k, \eta) F(t, k, \eta, y) + R_2(y', \underset{\sim}{\eta}) \tag{12}$$

where the coefficients can be calculated by

$$\alpha(t, k, \xi) = \frac{1}{k!} \int^{k} \bar{\alpha}_k(t, s_1, \ldots, s_k) \sum_{\pi} \phi_{\pi(\xi_1)}(s_1) \cdots \phi_{\pi(\xi_k)}(s_k) \, \mathrm{d}^k s$$

$$= \int^{k} \bar{\alpha}_k(t, s_1, \ldots, s_k) \prod_{j=1}^{k} \phi_{\xi_j}(s_j) \, \mathrm{d}^k s,$$

$$\beta(t, k, \eta) = \int^{k} \bar{\beta}_k(t, s_1, \ldots, s_k) \prod_{j=1}^{k} \phi_{\eta_j}(s_j) \, \mathrm{d}^k s. \tag{13}$$

We stress that (12) are orthogonal expansions wrt μ_w. Not μ_y!

3. Projection for Fourier-Hermite Series representation

Consider the set of all the FH series indexed by $\underset{\sim}{\xi}$ and denote it by

$$H_n(t, \underset{\sim}{\xi}) := \left\{ \sum_{k=0}^{n} \sum_{\xi \in \xi_k} a(t, k, \xi) F(t, k, \xi, y) \middle| a(t, k, \xi) \text{ are real numbers} \right\}.$$

We will call an element of $H_n(t, \underset{\sim}{\xi})$ the $\underset{\sim}{\xi}$-optimal estimate of $\Phi(x)$ and denote it by $\underset{\sim}{\xi}(\Phi) := \underset{\sim}{\xi}_t(\Phi(x))$, if

$$E(\underset{\sim}{\xi}_t(\Phi(x)) - \Phi(x))^2 \leq E(z - \Phi(x))^2, \quad \forall z \in H_n(t, \xi).$$

As

$$\sum_{\xi \in \xi_k} a(t, k, \xi) \prod_{j=1}^{k} \phi_{\xi_j}(t, s_j) \in L^2([0, t]^k), \quad k = 1, \ldots, n,$$

we see that $H_n(t, \underset{\sim}{\xi})$ is a closed Hilbert subspace of $H_n(t)$:

$$H_n(t, \underset{\sim}{\xi}) \subset H_n(t)$$

$$:= \left\{ \sum_{k=0}^{m} \int^{k} a_k(t, s_1, \ldots, s_k) \, \mathrm{d}^k y \middle| a_k \in L^2([0, t]^k), k = 0, \ldots, n \right\}.$$

Let the norm of the Hilbert space H of all finite variance random variables z be denoted by $\|z\| = (Ez^2)^{1/2}$.

It follows immediately that

(i) $\underset{\sim}{\xi}_t(\Phi)$ is unique;

(ii) An element z_n of $H_n(t, \underset{\sim}{\xi})$ is $\underset{\sim}{\xi}$-optimal, if $E[(\Phi - z_n)z] = 0, \forall z \in H_n(t, \underset{\sim}{\xi})$, i.e., z_n is the projection of $\Phi(x)$ onto $H_n(t, \underset{\sim}{\xi})$.

(iii) *If* $\underset{\sim}{\xi} \subset \underset{\sim}{\eta}$, then $\|\Phi - \underset{\sim}{\eta}_t(\Phi)\|^2 = \|\Phi - \underset{\sim}{\xi}_t(\Phi)\|^2 - \|\underset{\sim}{\xi}_t(\Phi) - \underset{\sim}{\eta}_t(\Phi)\|^2$.

The projection theorem to be developed provides a simple method of projecting $\Phi(x)$ onto $H_n(t, \underset{\sim}{\xi})$. First, we prove a lemma.

Lemma 3.1. *Consider the Fourier-Hermite series representation* (11) *of* $\sigma_t(\Phi)$. *For every* $z(y) \in H_n(t, \underset{\sim}{\xi})$,

$$E\left[\left(\Phi(x)(\sigma(1))(w) - \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \alpha(t, k, \xi) F(t, k, \xi, w)\right) z(w)\right] = 0. \tag{14}$$

Proof. LHS of (14)

$$= E\left[\left(\Phi(x)(\sigma_t(1))(w) - \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \alpha(t, k, \xi) F(t, k, \xi, w) - R(w', \underset{\sim}{\xi})\right) z(w)\right]$$

$$= E[(\Phi(x)(\sigma_t(1))(w) - (\sigma_t(\Phi))(w)) z(w)]$$

$$= E\left[\left(\frac{\Phi(x)(\sigma_t(1))(w) - (\sigma_t(\Phi))(w)}{(\sigma_t(1))(w)}\right) z(w) \frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_w}(w)\right]$$

$$= E[(\Phi(x) - \hat{\Phi}(x)) z(y)]$$

$$= 0.$$

Theorem 3.1. *The functional* $z_n(y) \in H_n(t, \underset{\sim}{\xi})$ *is the* $\underset{\sim}{\xi}$*-optimal estimate of* $\Phi(x)$ *if and only if*

$$D(y) := z_n(y)\sigma_t(1) - \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \alpha(t, k, \xi) F(t, k, \xi, y) \tag{15}$$

is orthogonal to $H_n(t, \underset{\sim}{\xi})$, *w.r.t.* μ_w, *i.e., for every* $z(y) \in H_n(t, \underset{\sim}{\xi})$,

$$E(D(w) z(w)) = 0. \tag{16}$$

Proof. (if): For every $z(y) \in H_n(t, \xi)$,

$$E[(\Phi(x) - z_n(y)) z(y)] = E[(\Phi(x)(\sigma_t(1))(w) - z_n(w)(\sigma_t(1))(w)) z(w)]$$

$$= E\left[\left(\Phi(x)(\sigma_t(1))(w) - \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \alpha(t, k, \xi) F(t, k, \xi, w) - D(w)\right) z(w)\right]$$

$$= 0.$$

The second equality follows from (15) and the third is a consequence of Lemma 3.1. and (16).

(only if): For every $z(y) \in H_n(t, \underset{\sim}{\xi})$,

$$E(D(w)z(w)) = E\left[\left(z_n(w)(\sigma_t(1))(w) - \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \alpha(t, k, \xi) F(t, k, \xi, w)\right) z(w)\right]$$
$$= E[(z_n(w)(\sigma_t(1))(w) - \Phi(x)(\sigma_t(1))(w))z(w)]$$
$$= E[(z_n(y) - \Phi(x))z(y)]$$
$$= 0.$$

The second equality results from Lemma 3.1.

4. The $\underset{\sim}{\xi}$-optimal estimation

Given an index set $\underset{\sim}{\xi}$, let the $\underset{\sim}{\xi}$-optimal estimate $\underset{\sim}{\xi}(\Phi(x))$ be written as

$$\underset{\sim}{\xi}(\Phi) = b(t, 0) + \sum_{k=1}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} b(t, k, \xi) F(t, k, \xi, y) \tag{17}$$

where $b(t, k, \xi)$ for $\xi \in \underset{\sim}{\xi}_k$ and $k = 1, \ldots, n$, are unknown functions of t to be determined.

Recall that the Fourier-Hermite (FH) series (12) of $\sigma_t(\Phi)$ and $\sigma_t(1)$ can be calculated from their iterated series (11) for any given index sets $\underset{\sim}{\xi}$ and $\underset{\sim}{\eta}$. We will, in this section, apply Theorem 3.1 to project $\hat{\Phi}(x) = \sigma_t(\Phi)/\sigma_t(1)$ onto $H_n(t, \underset{\sim}{\xi})$. The projection can be obtained by considering all the FH components of $\sigma(1)$ that affect each such ζ-component of the product $\xi_t(\Phi)\sigma_t(1)$ that $\zeta \in \underset{\sim}{\xi}$. The index set for all those FH components of $\sigma_t(1)$ will be denoted by $\underset{\sim}{\eta}$. Its determination necessitates the following lemmas:

Lemma 4.1. *The product of two FH functionals can be expressed as follows*:

$$F(t, k_1, \xi, y) F(t, k_2, \eta, y)$$
$$= \sum_{k=0}^{\min(k_1,k_2)} \frac{1}{k!} \int^{k_1+k_2-2k} \sum_{p_1} \sum_{p_2} \delta(p_1(\xi_1), p_2(\eta_1)) \cdots \delta(p_1(\xi_k), p_2(\eta_k))$$
$$\cdot \sum_{\pi} \phi_{p_1(\xi_{k+1})}(s_{\pi(1)}) \cdots \phi_{p_1(\xi_{k_1})}(s_{\pi(k_1-k)})$$
$$\cdot \phi_{p_2(\eta_{k+1})}(s_{\pi(k_1-k+1)}) \cdots \phi_{p_2(\eta_{k_2})}(s_{\pi(k_1+k_2-2k)}) \, \mathrm{d}^{k_1+k_2-2k} y, \tag{18}$$

$$\delta(i, j) := \begin{cases} 1 & \textit{for } i = j, \\ 0 & \textit{for } i \neq j, \end{cases}$$

where p_1 and p_2 are permutations of $(\xi_1, \ldots, \xi_{k_1})$ and $(\eta_1, \ldots, \eta_{k_2})$ respectively, and $\sum_{p_1} \sum_{p_2}$ is the summation over all such permutations.

Proof. An immediate consequence of Theorem 2.1.

From this expression, we see in order that the product (18) have a ζ-component, the vector η must contain all the components in ξ and ζ other than those that ξ and ζ have in common. Here repeated components are regarded as distinct one's. For instance, if $\xi=[1,2,2,3]$ and $\zeta=[2,3,4]$, then η must contain $\{1,2,4\}$. Furthermore, the vector η may contain any number of the common components of ξ and ζ, but they have to be included in duplicate. In the above example, the common components are $\{2,3\}$ and the set of all η's that make (18) have a ζ-component is $\{[1,2,4],[1,2,2,2,4],[1,2,3,3,4],[1,2,2,2,3,3,4]\}$.

There is a bijective relationship between index vectors and products of ϕ's, namely $[i_1,\ldots,i_k]\leftrightarrow\phi_{i_1}\cdots\phi_{i_k}$, where ϕ's are regarded as distinct variables and $i_1\leqslant\cdots\leqslant i_k$. Let $G:=GCD(\phi_{\xi_1}\cdots\phi_{\xi_{k_1}},\phi_{\zeta_1}\cdots\phi_{\zeta_{k_3}})$, $F:=\phi_{\xi_1}\cdots\phi_{\xi_{k_1}}/G$, and $H:=\phi_{\zeta_1}\cdots\phi_{\zeta_{k_3}}/G$. If the power of G is l, then there are less than or equal to 2^l *different* factors (including 1) of it. Taking the square of each factor and multiplying it to F · H yields the set $\underset{\sim}{\eta}(\xi,\zeta)$ of all the η that make (18) have a ζ-component.

When $\eta\in\underset{\sim}{\eta}(\xi,\zeta)$, we see that

$$\begin{aligned}&\text{the } \zeta\text{-component of (18)}\\&\quad=C(\xi,\eta,\zeta)F(t,k(\zeta),\zeta,y)/[k(\xi)+k(\eta)-k(\zeta))/2]!\end{aligned}\tag{19}$$

where $k(\xi)$, $k(\eta)$, and $k(\zeta)$ are the dimension of ξ, η, and ζ, respectively, and the constant $C(\xi,\eta,\zeta)$ results from the permutations p_1 and p_2 in (18). To see how $C(\xi,\eta,\zeta)$ can be determined, let us first consider the example: $\xi=[1,2_1,2_2,3]$, $\eta=[1,2_3,2_4]$, and $\zeta=[2,2,3]$. The subscripts of 2 are used to distinguish different permutations. Simple substitution yields that

$$\begin{aligned}C(\xi,\eta,\zeta)&=\delta(1,1)\delta(2_1,2_3)+\delta(2_1,2_3)\delta(1,1)+\delta(1,1)\delta(2_1,2_4)\\&\quad+\delta(2_1,2_4)\delta(1,1)+\delta(1,1)\delta(2_2,2_3)+\delta(2_2,2_3)\delta(1,1)\\&\quad+\delta(1,1)\delta(2_2,2_4)+\delta(2_2,2_4)\delta(1,1)\\&=(\#\text{ ways that }\{1,2\}\text{ can be drawn from }\xi)\\&\quad\cdot(\#\text{ ways that }\{1,2\}\text{ can be drawn from }\eta)\\&\quad\cdot(\#\text{ permutations of }\{1,2_1\}\text{ or }\{1,2_2\})\\&\quad\cdot(\#\text{ ways that }\{1,2_3\}\text{ or }\{1,2_4\}\text{ can match a}\\&\qquad\text{permutation of }\{1,2_4\})\\&=2\cdot2\cdot2\cdot1=8.\end{aligned}$$

Recall the bijective relationship: $[i_1,\ldots,i_k]\leftrightarrow\phi_{i_1}\cdots\phi_{i_k}$. Denote by γ the index vector associated with $(\phi_{\xi_1}\cdots\phi_{\xi_{k(\xi)}}\phi_{\eta_1}\cdots\phi_{\eta_{k(\eta)}}(\phi_{\zeta_1}\cdots\phi_{\zeta_{k(\zeta)}})^{-1})^{1/2}$. A general formula for $C(\xi,\eta,\zeta)$ is then the following:

$$\begin{aligned}C(\xi,\eta,\zeta)=(&\#\text{ ways that the components of }\gamma\text{ can be drawn from those}\\&\text{of }\xi\text{ with repeated components subscripted for distinction})\end{aligned}$$

$\cdot$ (# ways that the components of γ can be drawn from those of η) $\cdot$ ((# components of γ)!)

$\cdot$ ((# components of γ)!/(# permutations of the components of γ with subscripts removed)).

Remark. Denote the index set of the RHS of (18) by $\underset{\sim}{\zeta}(\xi, \eta)$. Some reflection shows that $\underset{\sim}{\zeta}(\xi, \eta)$ can be constructed from ξ and η in exactly the same way as $\underset{\sim}{\eta}(\xi, \zeta)$ is constructed from ξ and ζ. Combining this observation and (19), Lemma 4.1 can be converted into a product-sum formula for FH functionals:

$$F(t, k(\xi), \xi, y)F(t, k(\eta), \eta, y) = \sum_{k=|k(\xi)-k(\eta)|}^{k(\xi)+k(\eta)} \sum_{\zeta\in\underset{\sim}{\zeta}_k(\xi,\eta)} \frac{C(\xi, \eta, \zeta)}{[(k(\xi)+k(\eta)-k)/2]!} F(t, k, \zeta, y) \tag{20}$$

where $\underset{\sim}{\zeta}_k(\xi, \eta)$ is the collection of all k-tuples in $\underset{\sim}{\zeta}(\xi, \eta)$. We note that $\underset{\sim}{\zeta}_k(\xi, \eta)$ is empty when $(k(\xi)+k(\eta)-k)/2$ is not an integer.

Lemma 4.2. *A product-to-sum formula for the finite FH series is*

$$\begin{aligned}&\left(\sum_{k=0}^{n} \sum_{\xi\in\underset{\sim}{\xi}_k} \gamma_1(t, k, \xi)F(t, k, \xi, y)\right)\left(\sum_{k=0}^{n} \sum_{\eta\in\underset{\sim}{\eta}_k} \gamma_2(t, k, \eta)F(t, k, \eta, y)\right)\\ &\quad= \sum_{\zeta\in\underset{\sim}{\zeta}(\underset{\sim}{\xi},\underset{\sim}{\eta})}\left[\sum_{k=0} \sum_{\xi\in\underset{\sim}{\xi}_k} \sum_{\eta\in\underset{\sim}{\eta}\cap\underset{\sim}{\eta}(\xi,\zeta)} \frac{C(\xi, \eta, \zeta)\gamma_1(t, k, \xi)\gamma_2(t, k(\eta), \eta)}{[(k+k(\eta)-k(\zeta))/2]}\right]\\ &\qquad\cdot F(t, k(\zeta), \zeta, y),\end{aligned} \tag{21}$$

$$\underset{\sim}{\zeta}(\underset{\sim}{\xi}, \underset{\sim}{\eta}) := \bigcup_{\xi\in\underset{\sim}{\xi}} \bigcup_{\eta\in\underset{\sim}{\eta}} \zeta(\xi, \eta).$$

Proof. We take the ξth term $\gamma_1 F$ and an index vector ζ, and consider all the index vectors of $\underset{\sim}{\eta}$ that belong to the index set $\underset{\sim}{\eta}(\xi, \zeta)$. Multiplying $\gamma_1 F$ to each $\gamma_2 F$ indexed by an element of $\underset{\sim}{\eta} \cap \underset{\sim}{\eta}(\xi, \zeta)$ yields the summation $\sum_{\eta\in\underset{\sim}{\eta}\cap\underset{\sim}{\eta}(\xi,\zeta)}$. We first sum over all $\gamma_1 F$ terms to get $\sum_{k=0}\sum_{\xi\in\underset{\sim}{\xi}_k}$ and then sum over all the ζ-components that may possibly result from the multiplications to get $\sum_{\zeta\in\underset{\sim}{\zeta}(\underset{\sim}{\xi},\underset{\sim}{\eta})}$.

Now we consider the product

$$\left(\sum_{k=0}^{n} \sum_{\xi\in\xi_k} b(t, k, u)F(t, k, \xi, y)\right)\left(\sum_{k=0}^{2n} \sum_{\eta\in\underset{\sim}{\eta}_k(\xi,\zeta)} \beta(t, k, \eta)F(t, k, \eta, y)\right).$$

Applying Lemma 4.2 and setting the ζ-components of $\xi_t(\Phi)\sigma_t(1)$ and $\sigma_t(\Phi)$ equal, we obtain

$$\sum_{k=0}^{n} \sum_{\xi\in\underset{\sim}{\xi}_k} \sum_{\eta\in\underset{\sim}{\eta}(\xi,\zeta)} \frac{C(\xi, \eta, \zeta)b(t, k, \xi)\beta(t, k(\eta), \eta)}{((k+k(\eta)-k(\zeta))/2)!} F(t, k(\zeta), \zeta, y) = \alpha(t, k(\zeta), \zeta)F(t, k(\zeta), \zeta, y).$$

Hence, for every $\zeta \in \underset{\sim}{\xi}_k$, $k = 0, \ldots, n$,

$$\sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \left(\sum_{\eta \in \underset{\sim}{\eta}(\xi,\zeta)} \frac{C(\xi, \eta, \zeta)\beta(t, k(\eta), \eta)}{[(k + k(\eta) - k(\zeta))/2]!} b(t, k, \xi) \right) = \alpha(t, k(\zeta), \zeta). \tag{22}$$

We note that these form an algebraic system of simultaneous linear equations, in which the number of unknowns $b(t, k, \xi)$ is the same as that of the equations. The solvability of (22) is guaranteed by the obvious existence of the projection of $\Phi(x)$ onto $H_n(t, \underset{\sim}{\xi})$.

5. Finite Fourier expansion of $b(t, k, \xi)$

It was shown in Section 4 that the functions $b(t, k, \xi)$ can easily be determined at each t by solving a linear system of simultaneous equations (22). Although the solution of (22) can be carried out at a large (but finite) number of time points before the implementation of the estimator, the storage of all $b(t, k, \xi)$ is obvious difficult especially when the time interval $[0, T]$ over which the estimator will be applied is large.

Therefore we will, in this section, expand $b(t, k, \xi)$ in a finite Fourier series. By choosing an appropriate set of orthonormal functions on $[0, T]$, which can easily be stored or reproduced, the above difficulty can be removed.

The expansion can be obtained by minimizing the cumulative error $\rho(\underset{\sim}{\xi}_t(\Phi), \hat{\hat{\Phi}})$ over the constant coefficients of $b_j(k, \xi)$, where

$$\rho(\underset{\sim}{\xi}_t(\Phi), \hat{\hat{\Phi}}) := \int_0^T E(\underset{\sim}{\xi}_t(\Phi) - \hat{\hat{\Phi}})^2 \nu_t \, \mathrm{d}t,$$

$$\hat{\hat{\Phi}}(x) := \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} b'(t, k, \xi) F(t, k, \xi, y),$$

$$b'(t, k, \xi) :=: \sum_{j=0}^{j(k,\xi)} b_j(k, \xi)\psi_j(t). \tag{23}$$

Here $\{\psi_j, j = 1, 2, \ldots\}$ on $[0, T]$ is a CONS on $[0, T]$ w.r.t. the weight function ν.

We will now set about calculating $\rho(\xi_t(\Phi), \hat{\hat{\Phi}})$. Applying Lemma 4.2, we obtain

$$(\xi_t(\Phi) - \hat{\hat{\Phi}})^2 = \sum_{\zeta \in \underset{\sim}{\zeta}(\underset{\sim}{\xi}, \underset{\sim}{\xi})} f(t, k(\zeta), \zeta) F(t, k(\zeta), \zeta, y), \tag{24}$$

$$f(t, k(\zeta), \zeta) := \sum_{k=0}^{n} \sum_{\xi \in \underset{\sim}{\xi}_k} \sum_{\eta \in \underset{\sim}{\xi} \cap \underset{\sim}{\eta}(\xi,\zeta)} \frac{C(\xi, \eta, \zeta)\gamma(t, k, \xi)\gamma(t, k(\eta), \eta)}{[(k + k(\eta) - k(\zeta))/2]!},$$

$$\gamma(t, k, \xi) := b(t, k, \xi) - \sum_{j=0}^{j(k,\xi)} b_j(k, \xi)\psi_j(t).$$

Since the FH functionals are orthonormal and the Radon-Nikodym derivative $\mathrm{d}\mu_y/\mathrm{d}\mu_w$ is $\sigma_t(1)$, it is easy to see from (23) and (12) that

$$\begin{aligned} E(\xi_t(\Phi)-\hat{\hat{\Phi}})^2 &= E((\xi_t(\Phi))(y^t)-(\hat{\hat{\Phi}})(y^t))^2 \\ &= E[((\xi_t(\Phi))(w^t)-(\hat{\hat{\Phi}})(w^t))^2(\sigma_t(1))(w)] \\ &= \sum_{\underset{\sim}{\zeta}\in\underset{\sim}{\zeta}(\underset{\sim}{\xi},\underset{\sim}{\xi})} \beta(t,k(\zeta),\zeta) f(t,k(\zeta),\zeta). \end{aligned}$$

Expressing $\rho(\xi_t(\Phi),\hat{\hat{\Phi}})$ as a quadratic polynomial in the unknown constants $b_j(k,\xi)$, we obtain

$$\begin{aligned} \rho(\xi_t(\Phi),\hat{\hat{\Phi}}) = &\sum_G B_0 + \sum_G \sum_{j=0}^{j(k,\xi)} B_1 b_j(k,\xi) + \sum_G \sum_{j=0}^{j(k(\eta),\eta)} B_2 b_j(k(\eta),\eta) \\ &+ \sum_G \sum_{j_1=0}^{j(k,\xi)} \sum_{j_2=0}^{j(k(\eta),\eta)} B_3 b_{j_1}(k,\xi) b_{j_2}(k(\eta),\eta), \end{aligned} \tag{25}$$

$$\sum_G := \sum_{\underset{\sim}{\zeta}\in\underset{\sim}{\zeta}(\underset{\sim}{\xi},\underset{\sim}{\xi})} \sum_{k=0}^{n} \sum_{\xi\in\underset{\sim}{\xi}_k} \sum_{\eta\in\underset{\sim}{\xi}\cap\underset{\sim}{\eta}(\xi,\zeta)}, \qquad Q := \frac{C(\xi,\eta,\zeta)\beta(t,k(\zeta),\zeta)}{[(k+k(\eta)-k(\zeta))/2]!},$$

$$B_0 := \int_0^T Q b^2(t,k,\xi)\nu_t\,\mathrm{d}t, \qquad B_1 := -\int_0^T Q b(t,k(\eta),\eta)\psi_j(t)\nu_t\,\mathrm{d}t,$$

$$B_2 := -\int_0^T Q b(t,k,\xi)\psi_j(t)\nu_t\,\mathrm{d}t, \qquad B_3 := \int_0^T Q\psi_{j_1}(t)\psi_{j_2}(t)\nu_t\,\mathrm{d}t.$$

By minimizing the above quadratic function over b_j, we get a finite Fourier expansion (23) of $b(t,k,\xi)$.

6. Recursive algorithms

By solving the algebraic system (22) of simultaneous equations, the FH coefficients $b(t,k,\xi)$ can be determined as functions of time. If the storage or the on-line calculation of these coefficients poses a difficult implementation problem, we can take a step further in approximating the minimum variance estimate. We can use the approximate estimator $\hat{\Phi}(x)$, for which only a finite number of constants $b_j(k,\xi)$ need to be stored. It seems always possible to choose a CONS on $[0,T]$ wrt the weight function ν of which the elements ψ_j can be recursively calculated in time. We should stress here that the approximate estimator $\hat{\hat{\Phi}}(x)$ so obtained is optimal for its given structure:

$$\hat{\hat{\Phi}}(x) = \sum_{k=0}^{m} \sum_{\xi\in\underset{\sim}{\xi}_k} \sum_{j=0}^{j(k,\xi)} b_j(k,\xi)\psi_j(t)F(t,k,\xi,y). \tag{26}$$

More specifically, it can be easily proven that the constant coefficients $b_j(k,\xi)$ obtained by minimizing the quadratic function (25) minimizes $\int_0^T E(\Phi-\hat{\hat{\Phi}})^2\nu_t\,\mathrm{d}t$ as well.

It is known [2, 4] that a FH functional $F(t, k, \xi, y) := G_k(t, \prod_{j=1}^{k} \phi_{\xi_j}, y)$ can be expressed as a product of the Hermite polynomials of the Wiener integrals $\int_0^t \phi_{\xi_j}(s)\, \mathrm{d}y_s$: Define the symbol $\phi^{(i)}$ by $G_k(t, \phi^{(i)}, y) := G_k(t, \prod_{j=1}^{i} \phi(s_j), y)$. If ϕ_{λ_i} appears k_i times in $\prod_{j=1}^{k} \phi_{\xi_j}$ for $i = 1, \ldots, q$, then $\prod_{j=1}^{k} \phi_{\xi_j} = \prod_{i=1}^{q} \phi_{\lambda_i}^{(k_i)}$ and

$$F(t, k, \xi, y) = \prod_{i=1}^{q} H_{k_i}\left(\int_0^t \phi_{\lambda_i}\, \mathrm{d}y\right). \tag{27}$$

Thus we have reduced the ξ-optimal estimation (17) to the calculation of the Wiener integrals $\int_0^t \phi_{\lambda_i}\, \mathrm{d}y$. Let us recall from (9) that ϕ_{λ_i} are orthonormal functions in s on $[0, t]$. They are hence functions $\phi_{\lambda_i}(t, s)$ of both t and s. Their depenedence on t has not been indicated so as to simplify the notation. Most standard orthonormal functions on $[0, t]$ such as the Legendre polynomials are not separable in t and s and a differential equation cannot be obtained by differentiating the Wiener integral $\int_0^t \phi_{\lambda_i}(t, s)\, \mathrm{d}y_s$.

If we are only interested in estimation over a finite time interval, the Legendre polynomials can still be used as the CONS as follows: Consider the time-scaled Legendre polynomials on $[0, t]$,

$$\phi_k(t, s) := \sqrt{\frac{2n+1}{t}}\, P_k\left(\frac{2s}{t} - 1\right), \qquad P_k(t) := \sum_{j=0}^{[k/2]} \frac{(-1)^j (2k-2j)!}{2^k (k-j)!\, j!\, (k-2j)!}\, t^{k-2j}.$$

The polynomial ϕ_k is of degree k and hence can be expressed as a linear combination of the Laguerre polynomials L_i of degree $i \leq k$ as follows:

$$\phi_k(t, s) =: \sum_{j=0}^{k} C_{kj}(t) L_j(s, a) = \sum_{j=0}^{k} C_{kj}(t)\, \mathrm{e}^{at} l_j(s, a),$$

$$L_j(s, a) := \sqrt{2a} \sum_{i=0}^{j} \frac{(-1)^i j!}{i! [(j-i)!]^2} (2as)^{j-i}, \qquad l_j(s, a) := L_j(s, a)\, \mathrm{e}^{-as}$$

where the functions $C_{ij}(t)$ can be easily determined.

Hence, a typical Wiener integral which appears in (27) can be written as

$$\int_0^t \phi_k(t, s)\, \mathrm{d}y_s = \sum_{j=0}^{k} C_{kj}(t)\, \mathrm{e}^{\gamma t} \int_0^t \mathrm{e}^{-\gamma t + as} l_j(s, a)\, \mathrm{d}y_s$$

where γ is a positive constant to be chosen.

Let us call the integral on the right side an element functional of the observation y^t and denote it by $u_j(t)$. Then it satisfies the linear differential equation

$$du_j(t) = -\gamma u_j(t)\, \mathrm{d}t + \mathrm{e}^{(a-\gamma)t} l_j(t, a)\, \mathrm{d}y_t \tag{28}$$

and the above Wiener integral can be expressed simply as

$$\int_0^t \phi_k(t, s)\, \mathrm{d}y_s = \sum_{j=0}^{k} C_{kj}(t)\, \mathrm{e}^{\gamma t} u_j(t). \tag{29}$$

Thus the Wiener integral (29) can be calculated recursively by using the differential equation (28). Unfortunately, while (28) is a stable linear equation for a positive

γ, the time-varying coefficients $c_{k_j}(t)\,e^{\gamma t}$ increases exponentially. If $\gamma = 0$, then (28) is unstable. In fact $e^{at}l_j(t, a)$ is a Laguerre polynomial of degree j, which increases exponentially. Therefore this method can be only used for the estimation over a finite time interval.

When the estimation interval is the infinite interval $[0, \infty)$, the following trick provides uş with a stable recursive algorithm:

Pick a CONS$\{\xi_1, \xi_2, \ldots\}$ on $[0, \infty)$ such that $\int_0^\infty \xi_k^2(s)\,ds < \infty$ for every $k = 1, 2, \ldots$, and the restrictions of ξ_k to $[0, t]$ are linearly independent on $[0, t]$. For example, the Laguerre functions l_k form such a CONS. By the Gram-Schmidt procedure, we construct a sequence of orthonormal functions $\{\phi_k(t, s), k = 1, 2, \ldots\}$ on $[0, t]$, i.e.,

$$\int_0^t \phi_j(t, s)\phi_k(t, s)\,ds = \delta_{jk}$$

for all positive integers j and k.

We note that the functions $\phi_k(t, s)$ so constructed form a CONS on $[0, t]$ and are linear combinations of ξ_k. More specifically, for each k,

$$\phi_k(t, s) = \sum_{j=1}^{k} c(t, k, j)\xi_j(s)$$

for some constants $c(t, k, j)$.

Now the Wiener integral $\int_0^t \phi_k(t, s)\,dy_s$ is recursively calculated as follows:

$$\int_0^t \phi_k(t, s)\,dy_s = \sum_{j=1}^{k} c(t, k, j)u_j(t), \qquad du_j(t) = \xi_j(t)\,dy_t.$$

It can easily be proven by induction that

$$\lim_{t\to\infty} c(t, k, j) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j < k, \end{cases}$$

and that the variance of $u_j(t)$ remains bounded if so does that of $y(t)$, as t approaches infinity.

References

[1] N. Wiener, Nonlinear Problems in Random Theory (The Technology Press, M.I.T. and J. Wiley and Sons Inc., NY, 1958).

[2] M. Schetzen, The Volterra and Wiener Theories of Nonlinear Systems (J. Wiley and Sons, Inc., NY, 1980).

[3] V. Volterra, Theory of Functional and of Integral and Integro-Differential Equations (Dover Puclications, Inc., NY, 1959).

[4] R.M. Cameron and W.T. Martin, The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals, Ann. of Math. 48 (1947) 359-389.

[5] J. Katzenelson and L.A. Gould, The design of nonlinear filters and control systems, Part I, Information and Control 5 (1962) 108-143.

[6] S.K. Mitter and D. Ocone, Multiple integral expansions for nonlinear filtering, Proceedings of the 18th IEEE Conference on Decision and Control (Ft. Lauderdale, Florida, 1979).
[7] D. Ocone, Topics in Nonlinear Filtering Theory, Ph.D. Thesis, Dept. of Math., M.I.T., 1980.
[8] D. Ocone, Multiple integral expansions for nonlinear filtering, to appear in Stochastics.
[9] K. Ito, Multiple Wiener integrals, J. Math. Soc. Japan 3 (1951) 157-169.
[10] J.T.-H. Lo and S.K. Ng, Optimal orthogonal expansion for estimation II: Signal in counting observations, in: R.S. Bucy and J.M.F. Moura, eds., Proc. 1982 NATO ASI on Nonlinear Stochastic Problems, (D. Reidel, 1983).
[11] R.S. Bucy and P.D. Joseph, Filtering for Stochastic Processes with Applications to Guidance (Interscience, NY, 1968).
[12] G. Kallianpur, Stochastic Filtering Theory (Springer-Verlag, NY, 1980).
[13] R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes I: General Theory (Springer-Verlag, NY, 1977).
[14] H.P. McKean, Jr., Stochastic Integrals (Academic Press, NY and London, 1969).
[15] I. Shigekawa, Derivatives of Wiener functionals and absolute continuity of induced measures, J. Math. Kyoto Univ. 20(2) (1980) 263-289.